IBM Developer
SKILLS NETWORK

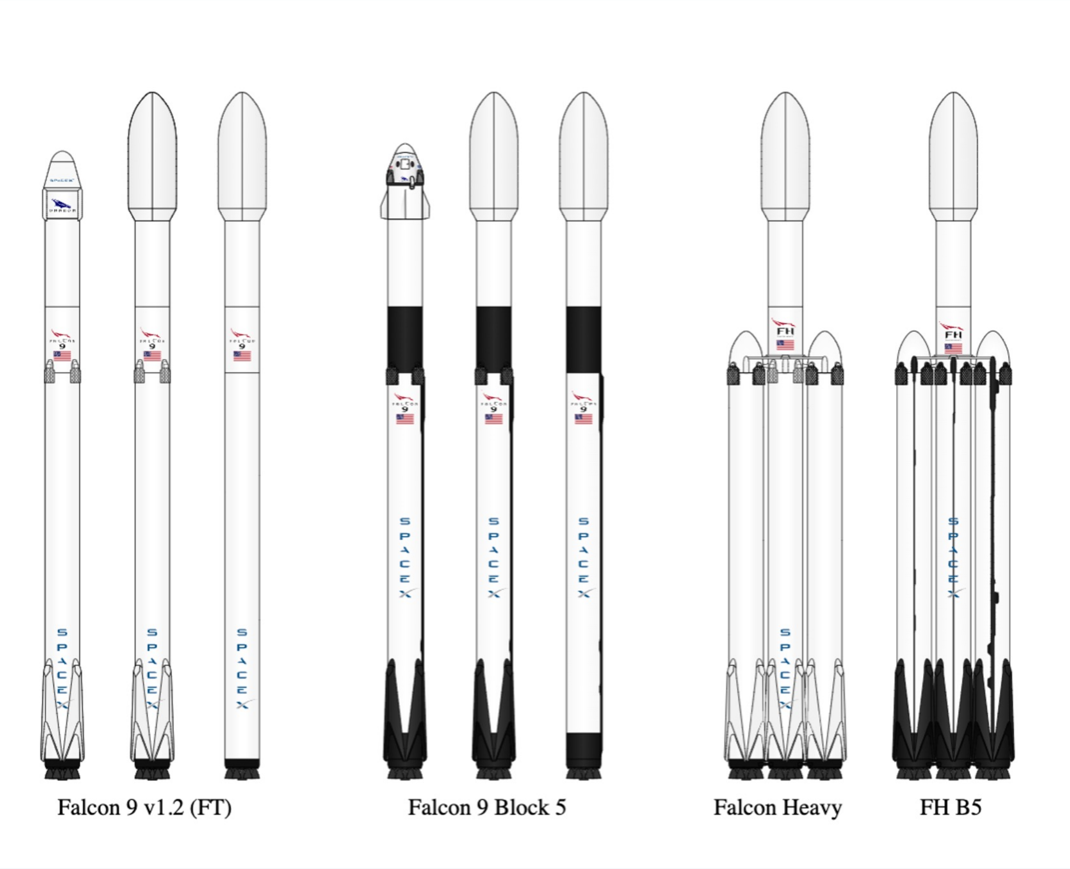
Winning Space Race with Data Science

Adriaan van Heerden
19.11.2024

Table of Contents

Executive Summary

- SpaceX advertises Falcon 9 rocket launches at a cost of $62m, compared to other providers that charge upwards of $165m
- SpaceX has a competitive advantage because it can reuse the first stage of the rocket
- If SpaceY can learn from SpaceX's experience and determine whether stage 1 rockets will land, then we can determine the cost of a launch and bid against SpaceX for future rocket launches
- Extensive data analysis and modelling were conducted to determine the factors for successful launches and for stage 1 rockets landing successfully
- The models produced in this report were able to identify success factors and predict stage 1 rocket landing success with an accuracy level of 83%



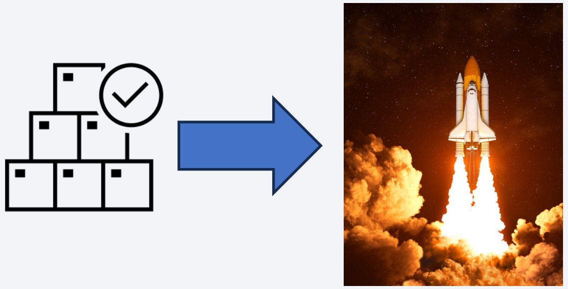
Introduction

Project background and context:

- SpaceX advertises Falcon 9 rocket launches at a cost of $62m (compared to other providers that charge upwards of $165m)
- SpaceX has a competitive advantage because it can reuse the first stage of the rocket
- If SpaceY can learn from SpaceX's experience and determine whether the first stage will land, then we can determine the cost of a launch and bid against SpaceX for rocket launches
- This project is designed to determine the factors for successful launches and for stage 1 rockets landing successfully

Questions that need answering:

- Which launch sites have the greatest success rates, and why?
- Which rockets have the greatest success rates?
- Can we predict with any certainty whether stage 1 rockets will land successfully?



Section 1

Methodology

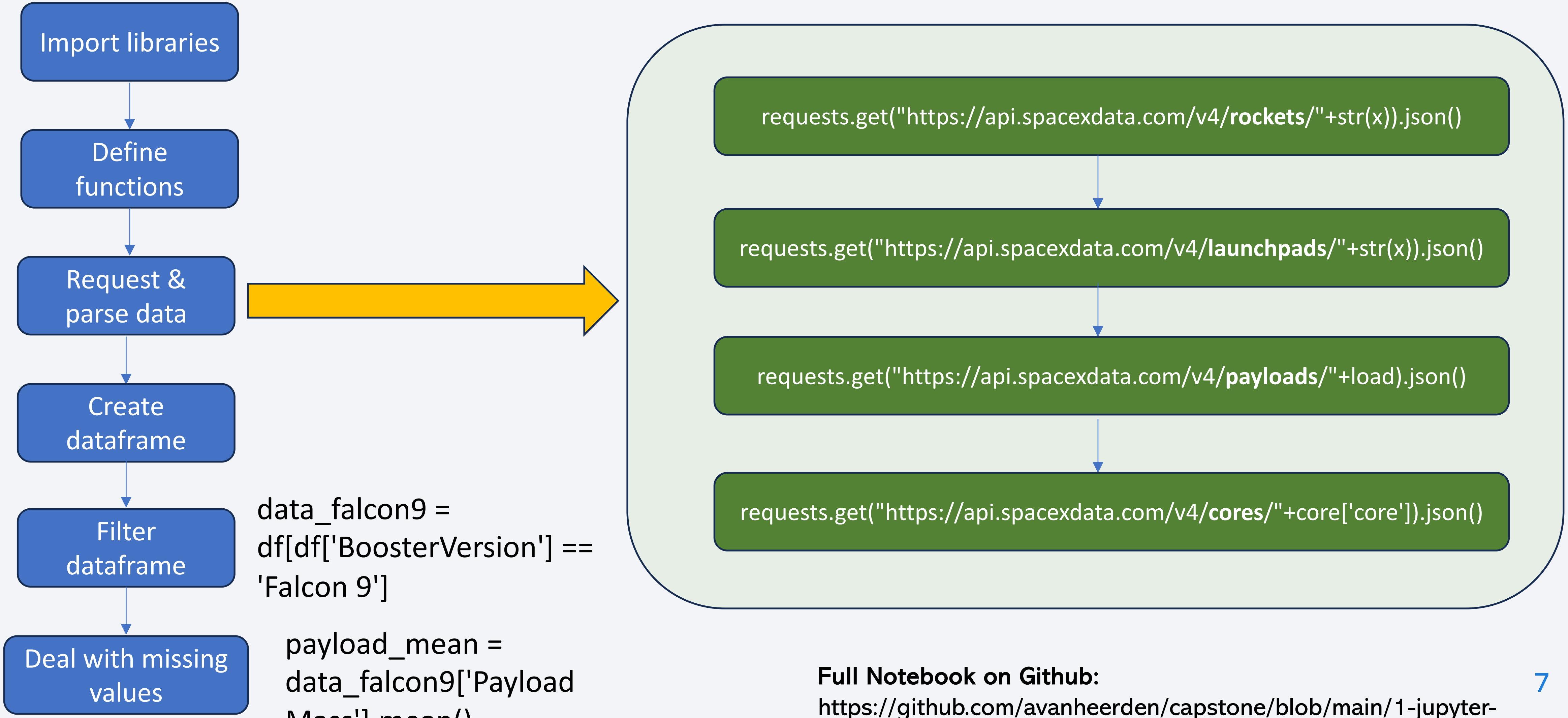
Methodology

Executive Summary

- Data collection methodology:
  - Relevant data were collected from Wikipedia using a BeautifulSoup webscraping object and directly from the Open Source REST API for SpaceX
- Perform data wrangling
  - Identify missing values in dataset, identify data types, map string data field ("Outcome") to a new "Class" field of 0 & 1 values for analysis
- Perform exploratory data analysis (EDA) using visualisation and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Four classification models were built, tuned and evaluated using SciKit Learn

Data Collection Method 1 – SpaceX API

https://api.spacexdata.com/v4/launches/past




```
data_falcon9 = df[df['BoosterVersion'] == 'Falcon 9']
```

```
payload_mean = data_falcon9['Payload Mass'].mean()
```

Full Notebook on Github:
https://github.com/avanheerden/capstone/blob/main/1-jupyter-labs-spacex-data-collection-api-v2.ipynb

Data Collection Method 2 – Webscraping

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Full Notebook on Github:
https://github.com/avanheerden/capstone/blob/main/2-jupyter-labs-webscraping.ipynb

Import libraries

Define functions

Request the Falcon9 Launch page from its URL

```
response = requests.get(static_url)
```

Create BeautifulSoup object

```
soup = BeautifulSoup(response.text, 'html.parser')
```

Extract all column/variable names from the HTML table header

```
html_tables = soup.find_all('table')
```

Create dataframe by parsing HTML tables

```
launch_dict= dict.fromkeys(column_names)
```

```
df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
```

Data Wrangling

Full Notebook on Github:
https://github.com/avanheerden/capstone/blob/main/3-labs-jupyter-spacex-Data%20wrangling-v2.ipynb

Import libraries

Combine gathered data into single dataset

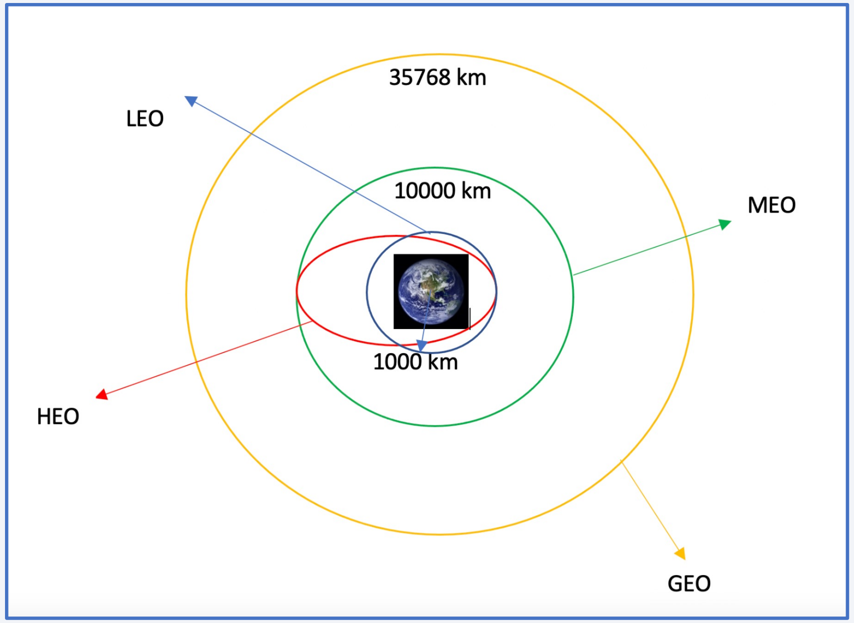
Load dataset into dataframe

Data Wrangling

Identify missing values

Calculate number of launches on each site

df['LaunchSite'].value_counts()



Calculate the number and occurrence of each orbit

df['Orbit'].value_counts()

Calculate the number and occurrences of mission outcomes for each orbit

landing_outcomes = df['Outcome'].value_counts().to_dict()

Create a landing outcome label from "Outcome" column

df['Class'] = df['Outcome'].apply(categorise_landing)

Class = Outcome

Class = 0; first stage booster did not land successfully
- None None; not attempted
- None ASDS; unable to be attempted due to launch failure
- False ASDS; drone ship landing failed
- False Ocean; ocean landing failed
- False RTLS; ground pad landing failed

Class = 1; first stage booster landed successfully
- True ASDS; drone ship landing succeeded
- True RTLS; ground pad landing succeeded
- True Ocean; ocean landing succeeded

Exploratory Data Analysis with SQL

Full Notebook on Github:
https://github.com/avanheerden/capstone/blob/main/4-jupyter-labs-eda-sql-coursera_sqllite.ipynb

- Loaded data into SQLite database:

```
df = pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv")
df.to_sql("SPACEXTBL", con, if_exists='replace', index=False,method="multi")
```

- Key queries performed:

Display the names of the unique launch sites in the space mission	%sql SELECT DISTINCT LANDING_OUTCOME FROM SPACEXTBL;
Display the total payload mass carried by boosters launched by NASA (CRS)	%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)'
Display average payload mass carried by booster version F9 v1.1	%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'
List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000	%sql SELECT PAYLOAD FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
List the names of the booster_versions which have carried the maximum payload mass	%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order	%SQL SELECT LANDING_OUTCOME, COUNT(*) AS COUNT_OUTCOMES FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING_OUTCOME ORDER BY COUNT_OUTCOMES DESC;

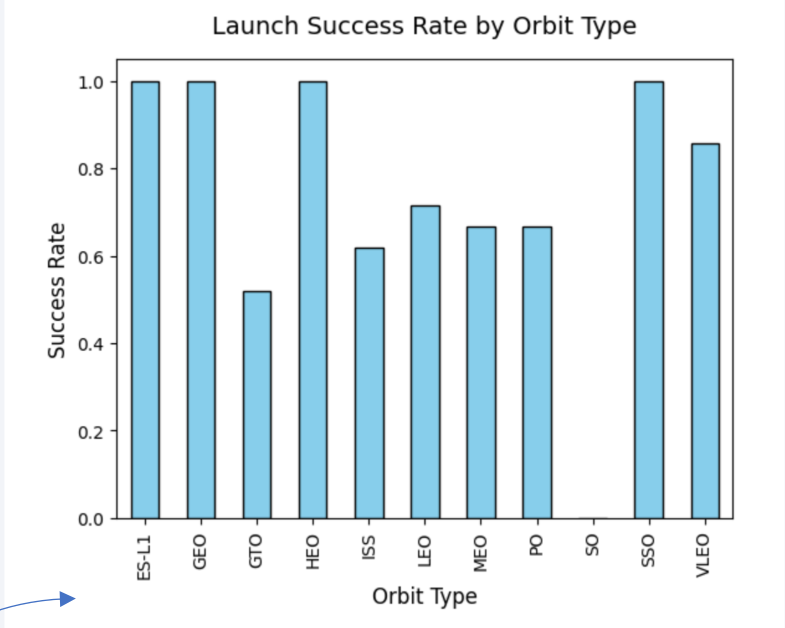
Exploratory Data Analysis with Data Visualisation

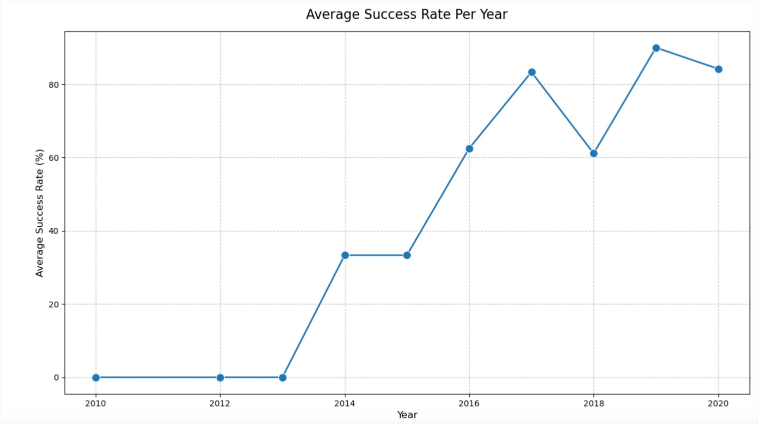
- Used the Seaborn library for data visualisation
- Used scatterplots to visualise relationships between:
  - Flight Number vs. Payload Mass
  - Flight Number vs. Launch Site
  - Payload Mass vs. Launch Site
  - Flight Number vs. Orbit Type
  - Payload Mass vs. Orbit Type

Scatterplots were used for the following reasons:

1. Direct Correlation Analysis: A scatterplot helps identify if there is any relationship between the variables plotted and their effect on the success or failure of a mission
2. Pattern Identification: A scatterplot helps to identify data "clusters" as well as any outliers in the data, which could signal the need for further analysis

- Used a bar chart to visualise launch success rate by orbit type: looking at the graph it is immediately evident that 4 orbits have the best success rate
- Used a line chart to visualise average success rate by year, to see if the success rate changed over time





Full Notebook on Github:
https://github.com/avanheerden/capstone/blob/main/5-edadataviz.ipynb

Exploring the Geography with a Folium Interactive Map

A Folium interactive map was generated in order to gain insights from the local geography of the launch sites and to visualise the most successful launch sites. The following tasks were completed:

- **TASK 1:** Mark all launch sites on a map after extracting latitude & longitude coordinates
- **TASK 2:** Mark the successful & failed launches for each site on the map
- **TASK 3:** Calculate the distances between a launch site to its proximities (nearest city, highway, etc.)

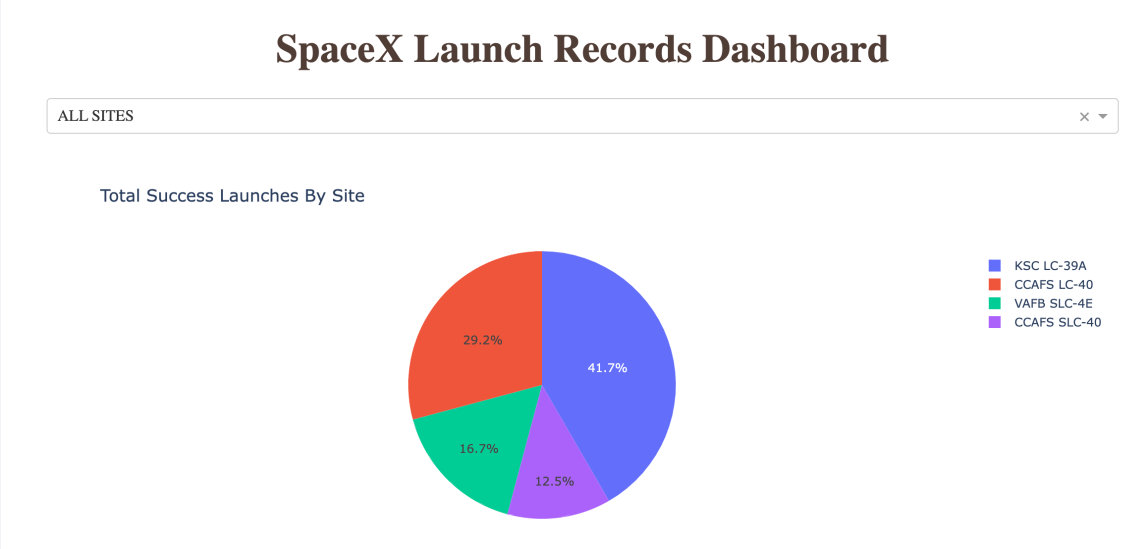
In order to complete these tasks, map objects such as markers, marker clusters, circles, lines and popup labels were created and added to the map to aid visualisation and analysis.

Full Notebook on Github:
https://github.com/avanheerden/capstone/blob/main/6-lab_jupyter_launch_site_location.ipynb

Interactive Plotly Dash Application

A Plotly Dash application was created for users to perform interactive visual analytics on SpaceX launch data in real time. The dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatterplot chart in order to answer the following 5 questions:

1. Which site has the largest number of successful launches?
2. Which site has the highest launch success rate?
3. Which payload range has the highest launch success rate?
4. Which payload range has the lowest launch success rate?
5. Which F9 Booster version has the highest launch success rate?



Full Notebook on Github:
https://github.com/avanheerden/capstone/blob/main/7-spacex_dash_app.py

Predictive Analysis (Classification)

Import libraries

Define function to create confusion matrix

from sklearn.metrics import confusion_matrix

Load SpaceX data obtained in previous stage

from sklearn.preprocessing import StandardScaler

Standardise data

X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)

Split data into training & test sets

Fit the training data to different predictive models

1. Logistic Regression
2. Support Vector Machine (SVM)
3. Decision Tree Classifier
4. K Nearest Neighbour (KNN)

Refine model parameters using GridSearchCV

Evaluate model accuracy to find best model

Four classification models were built, tested and refined in order to identify the best model for predicting future successful landings

Full Notebook on Github:
https://github.com/avanheerden/capstone/blob/main/8-SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Model Comparison Report:

Model	Test Accuracy
Logistic Regression	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

Results Obtained from EDA

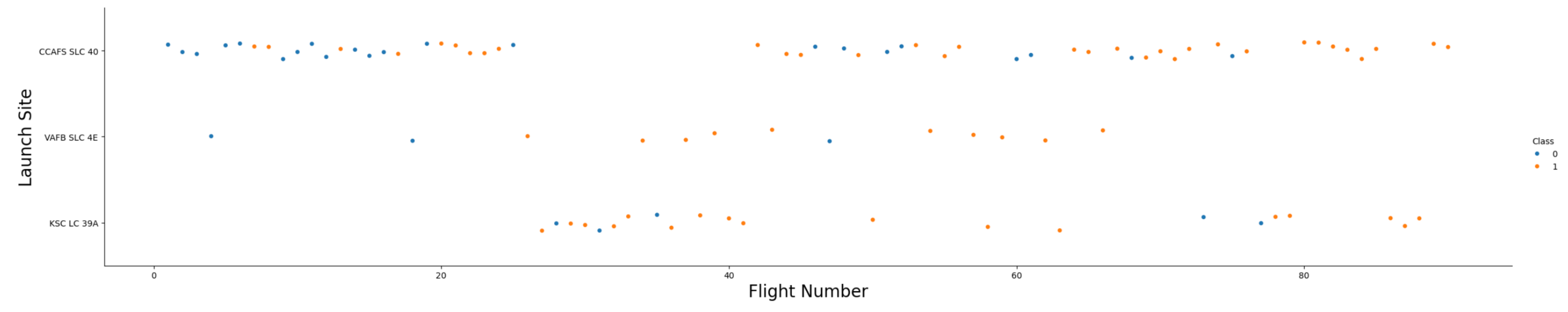
The sections below detail the exploratory data analysis results as follows:

- Scatterplots of relevant variables plotted against each other showing successful & failed landings
- Bar chart showing mission success rates for different orbits used
- Line chart showing mission success rate over time
- Results of SQL queries showing relevant key information
- Interactive Folium maps to explore the geographical features and success rates of launch sites
- Plotly Dash interactive dashboard to identify the launch sites with best success rates, optimal payloads and preferred orbits
- Comparison between four classification models to determine the model with the best accuracy for predicting future successful landings

Section 2

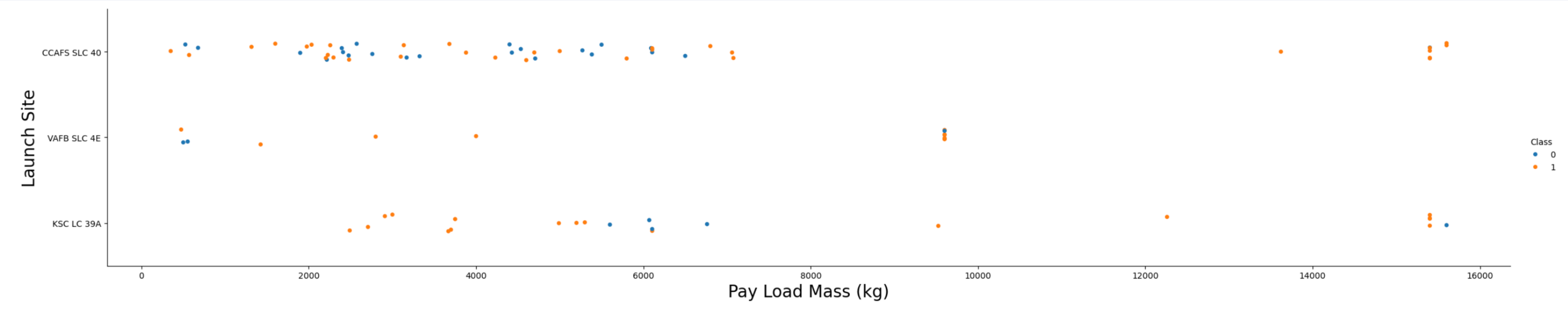
Insights drawn from EDA

Flight Number vs. Launch Site



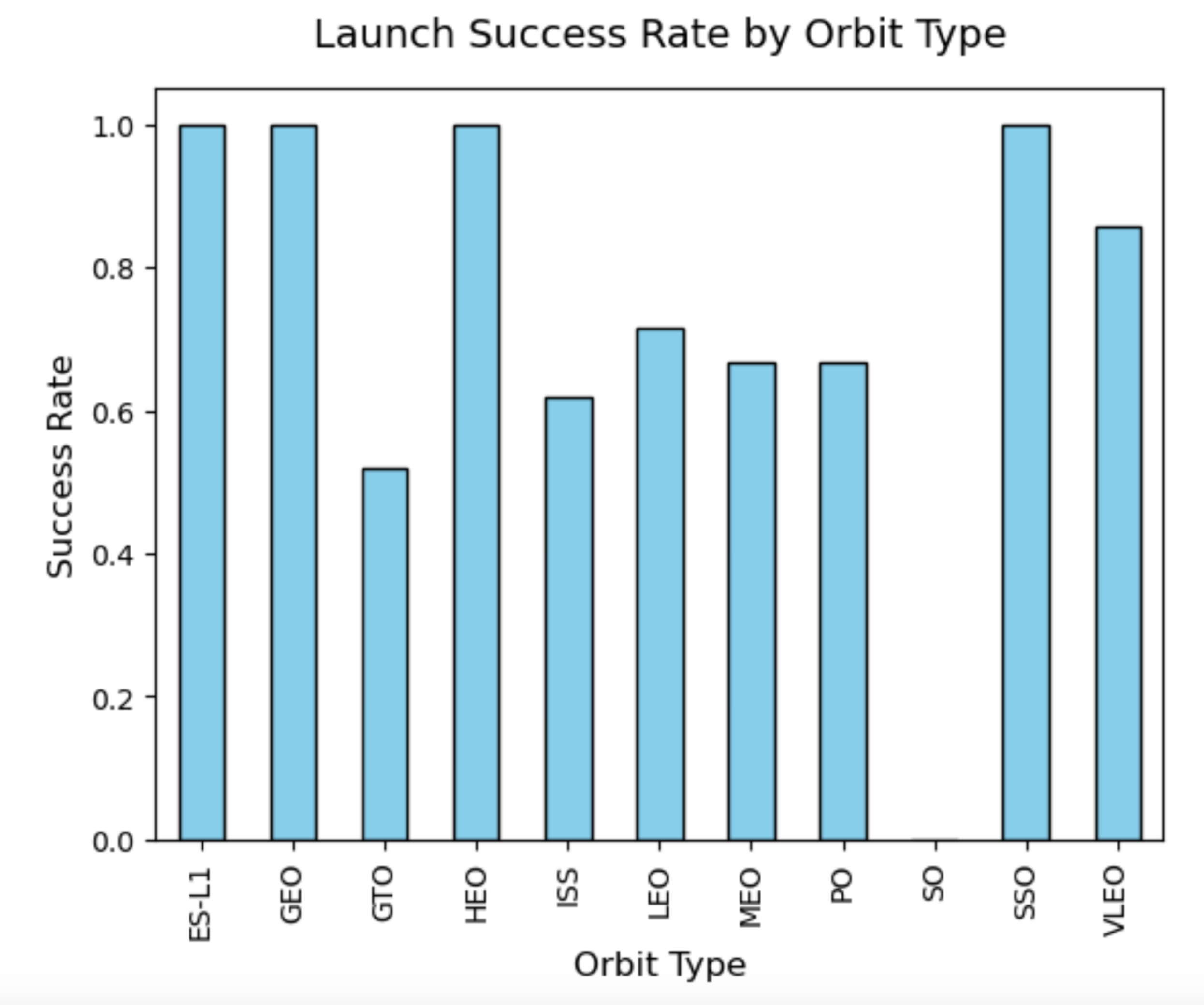
1. There are three launch sites shown:
    1. CCAFS SLC 40 [Cape Canaveral Air Force Station Space Launch Complex 40] (top)
    2. VAFB SLC 4E [Vandenberg Air Force Base Space Launch Complex 4E] (middle)
    3. KSC LC 39A [Kennedy Space Center Launch Complex 39A] (bottom)
2. CCAFS SLC 40 appears to be the most frequently used launch site, with the most data points
3. Pattern analysis:
    1. Early flights (lower flight numbers) seem to have more failures (0/blue dots = failure; 1/orange dots = success)
    2. Later flights (higher flight numbers) show more successes, suggesting improvement over time
    3. VAFB SLC 4E appears to have fewer launches overall compared to the other sites, with use of the site dropping off sometime after Flight Number 65

Payload vs. Launch Site



1. There are three launch sites shown:
    1. CCAFS SLC 40 [Cape Canaveral Air Force Station Space Launch Complex 40] (top)
    2. VAFB SLC 4E [Vandenberg Air Force Base Space Launch Complex 4E] (middle)
    3. KSC LC 39A [Kennedy Space Center Launch Complex 39A] (bottom)
2. Pattern analysis:
    1. There appears to be launches across a wide range of payload masses, from near 0 kg to about 16,000 kg
    2. CCAFS SLC 40 has the highest number of launches among the three sites
    3. The success rate (orange dots) seems relatively high across all sites
    4. Failed launches (blue dots) appear to be distributed across different payload masses, suggesting payload mass might not be a primary factor in launch success
    5. VAFB SLC 4E seems to have fewer launches overall compared to the other sites

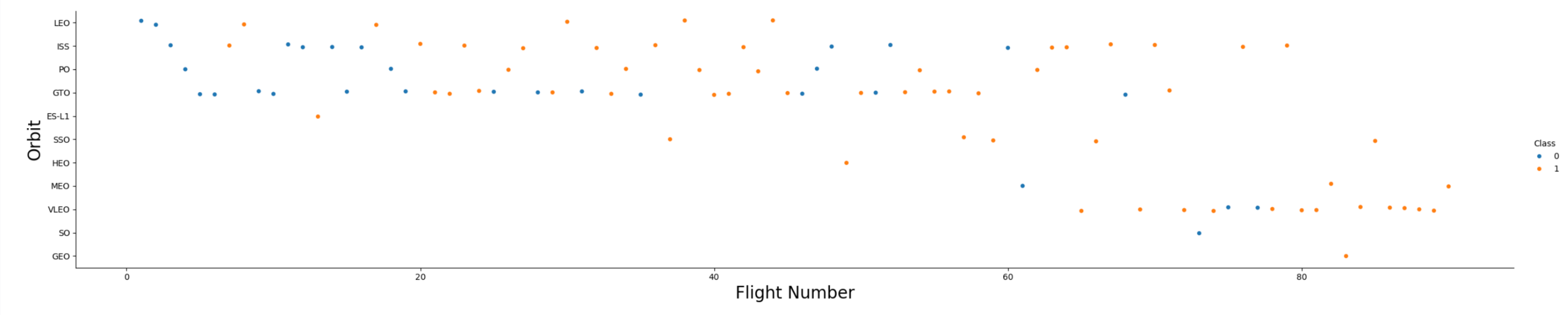
Success Rate vs. Orbit Type



Using a bar chart to visualise launch success rate by orbit type, it is immediately evident that 4 orbits have 100 % success rates: ES-11, GEO, HEO and SSO.

The use of the SO orbit was recorded once (for a failed mission), however the SO orbit should be classed with the SSO orbit (see Notebook 3), so this will have a slight impact on the overall SSO success rate.

Flight Number vs. Orbit Type



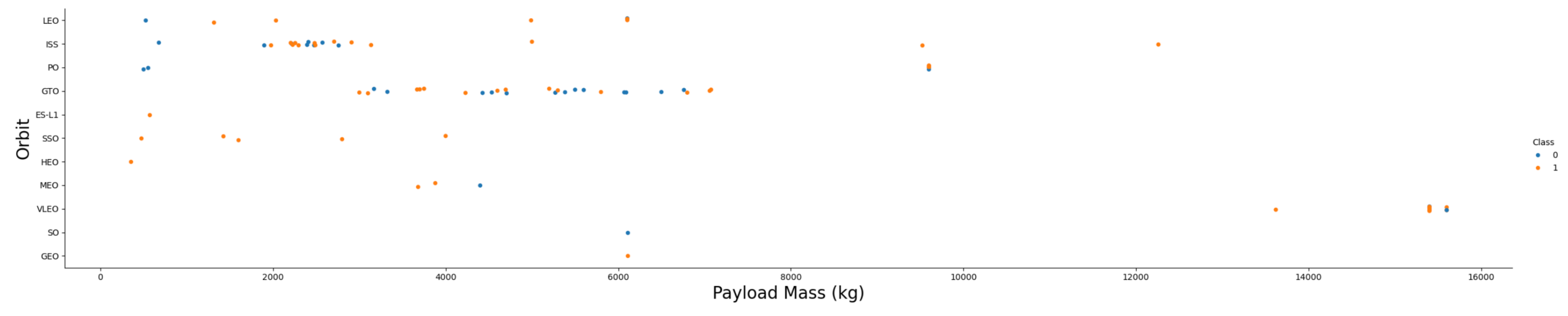
The different orbit types shown are:

1. LEO (Low Earth Orbit)
2. ISS (International Space Station orbit)
3. PO (Polar Orbit)
4. GTO (Geosynchronous Transfer Orbit)
5. ES-L1 (Earth-Sun Lagrange Point 1)
6. SSO/SO (Sun-Synchronous Orbit)
7. HEO (Highly Elliptical Orbit)
8. MEO (Medium Earth Orbit)
9. VLEO (Very Low Earth Orbit)
10. GEO (Geosynchronous Earth Orbit)

Pattern analysis:

1. Some orbit types (LEO, ISS, PO, GTO) are more frequently attempted
2. Some orbit types (like ES-L1, SSO, MEO, GEO) are less frequently attempted
3. There seems to be a learning curve - earlier flights (lower flight numbers) show more failures (blue dots)
4. Later flights (higher flight numbers) show a better success rate, suggesting improved reliability over time
5. The VLEO orbit is used more frequently in later flights, with good success rates

Payload vs. Orbit Type



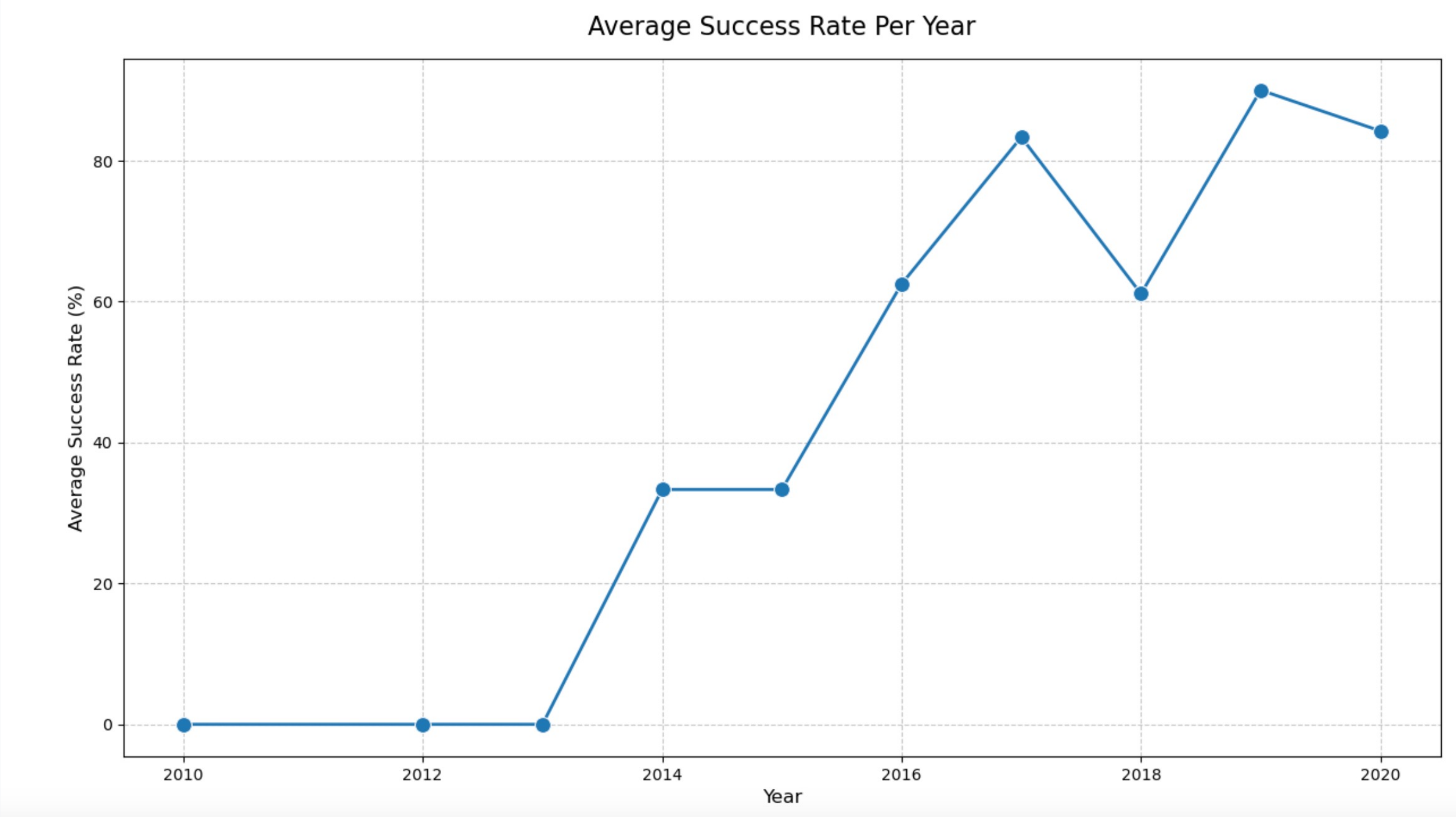
Pattern analysis:

1. Different orbit types tend to have characteristic payload mass ranges:
   - ISS missions cluster around 2,000-3,000 kg
   - GTO missions span a wider range, roughly 3,000-7,000 kg
   - LEO missions span a range of c. 1,000-6,000 kg
   - VLEO missions appear successful with heavier payloads (around 15,000 kg)

2. Payload distribution:
   - Most missions fall within 0-8,000 kg range, with a sharp drop off after c. 7,000 kg

3. Success/Failure patterns:
   - There appear to be more failures (blue dots) in the ISS and GTO orbits
   - Higher success rates (orange dots) are seen in the medium payload range (2,000-6,000 kg)
   - Even very heavy payloads (>10,000 kg) show good success rates
   - GTO missions show a mix of successes and failures across their payload range

Launch Success Yearly Trend

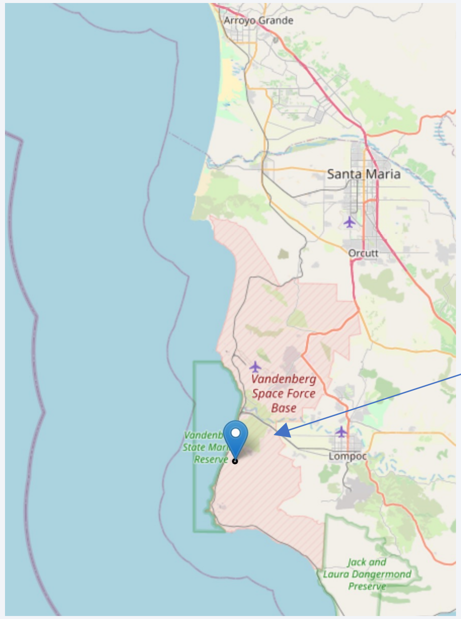


It can be observed from the line graph that the success rate kept increasing from 2013 until 2020, with three of the last four data points above 80%

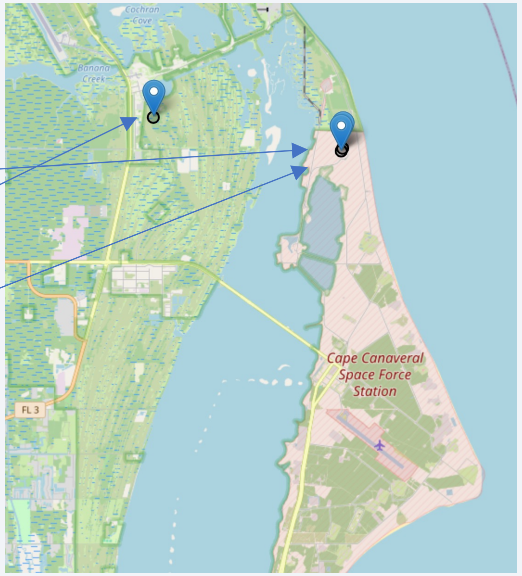
All Launch Site Names

- Imported dataset into SQLite database (SPACEXTBL)
- Ran SQL query to find the names of all the unique launch sites:

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTBL;
```



Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40



Launch Site Names Begin with 'CCA'

- Ran SQL query to find 5 records where launch site names begin with “CCA”:

```
%sql select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- It can be seen from the query result that the dataset has been filtered on the Launch_Site column as per the “where” clause

Total Payload Mass

- Ran SQL query to calculate the total payload carried by boosters from NASA:

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)'
```

SUM(PAYLOAD_MASS__KG_)
45596

- It can be seen from the query result that the total payload mass carried by boosters from NASA was about 46 tons (45,596 kg)

Average Payload Mass by F9 v1.1

- Ran SQL query to calculate the average payload mass carried by booster version F9 v1.1:

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'
```

AVG(PAYLOAD_MASS__KG_)
2928.4

- It can be seen from the query result that the average payload mass carried by booster F9 v1.1 was about 3 tons (2,928.4 kg)

First Successful Ground Landing Date

- Ran SQL query to find the date of the first successful landing outcome on ground pad:

```
%sql SELECT MIN(DATE) FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Success (ground pad)'
```

MIN(DATE)
2015-12-22

- It can be seen from the query result that the first successful landing was achieved on 22nd December 2015

Successful Drone Ship Landing with Payload between 4,000 and 6,000 kg

- Ran SQL query to list the names of boosters which have successfully landed on a drone ship and had a payload mass greater than 4,000 but less than 6,000 kg:

```
%sql SELECT Booster_Version, Payload, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 FT B1022	JCSAT-14	4696
F9 FT B1026	JCSAT-16	4600
F9 FT B1021.2	SES-10	5300
F9 FT B1031.2	SES-11 / EchoStar 105	5200

It can be seen from the query result that the F9 FT booster achieved four successful drone ship landings within the specified parameters

Total Number of Successful and Failure Mission Outcomes

- Ran SQL query to calculate the total number of successful and failed missions:

```
%sql SELECT MISSION_OUTCOME, COUNT(*) as total FROM SPACEXTBL GROUP BY MISSION_OUTCOME
```

Mission Outcome	Total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

It can be seen from the query result that there were 100 successful missions and one unsuccessful one

Boosters Carried Maximum Payload

- Ran SQL query to identify the name of the booster which carried the maximum payload mass:

```
%sql SELECT BOOSTER_VERSION, PAYLOAD_MASS__KG_ as Payload FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
ORDER BY BOOSTER_VERSION ASC
```

Booster_Version	Payload
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

It can be seen from the query result that the F9 B5 booster carried the maximum payload of 15,600 kg

2015 Launch Records

- Ran SQL query to identify the failed landing outcomes on drone ships, their booster versions, and launch site names in the year 2015:

```
%sql select substr(DATE, 6,2) as Month, BOOSTER_VERSION, LAUNCH_SITE, LANDING_OUTCOME from SPACEXTBL where (LANDING_OUTCOME = 'Failure (drone ship)') and (substr(DATE, 0,5)='2015');
```

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

It can be seen from the query result that there were two failed landings on drone ships in 2015 (in January and April), both of which were launched from the CCAFS LC-40 site using the F9 v1.1 booster

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Ran SQL query to rank the count of landing outcomes between 2010-06-04 and 2017-03-20, in descending order:

```
%%sql SELECT Landing_Outcome, count(*) as Total
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY Total DESC;
```

Landing_Outcome	Total
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Analysis on next slide

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

We can draw a few interesting conclusions from the query results on the previous slide:

1. Landing Attempts vs No Attempts:
- 21 missions had landing attempts
- 10 missions had no landing attempts
- This shows SpaceX attempted landings in roughly 68% of missions

2. Landing Success Rate:
- Successful landings: 8 total (5 drone ship + 3 ground pad)
- Failed landings: 8 total (5 drone ship + 2 parachute + 1 precluded)
- Controlled/Uncontrolled ocean landings: 5 total
- The success rate for attempted landings is about 38% (8/21)

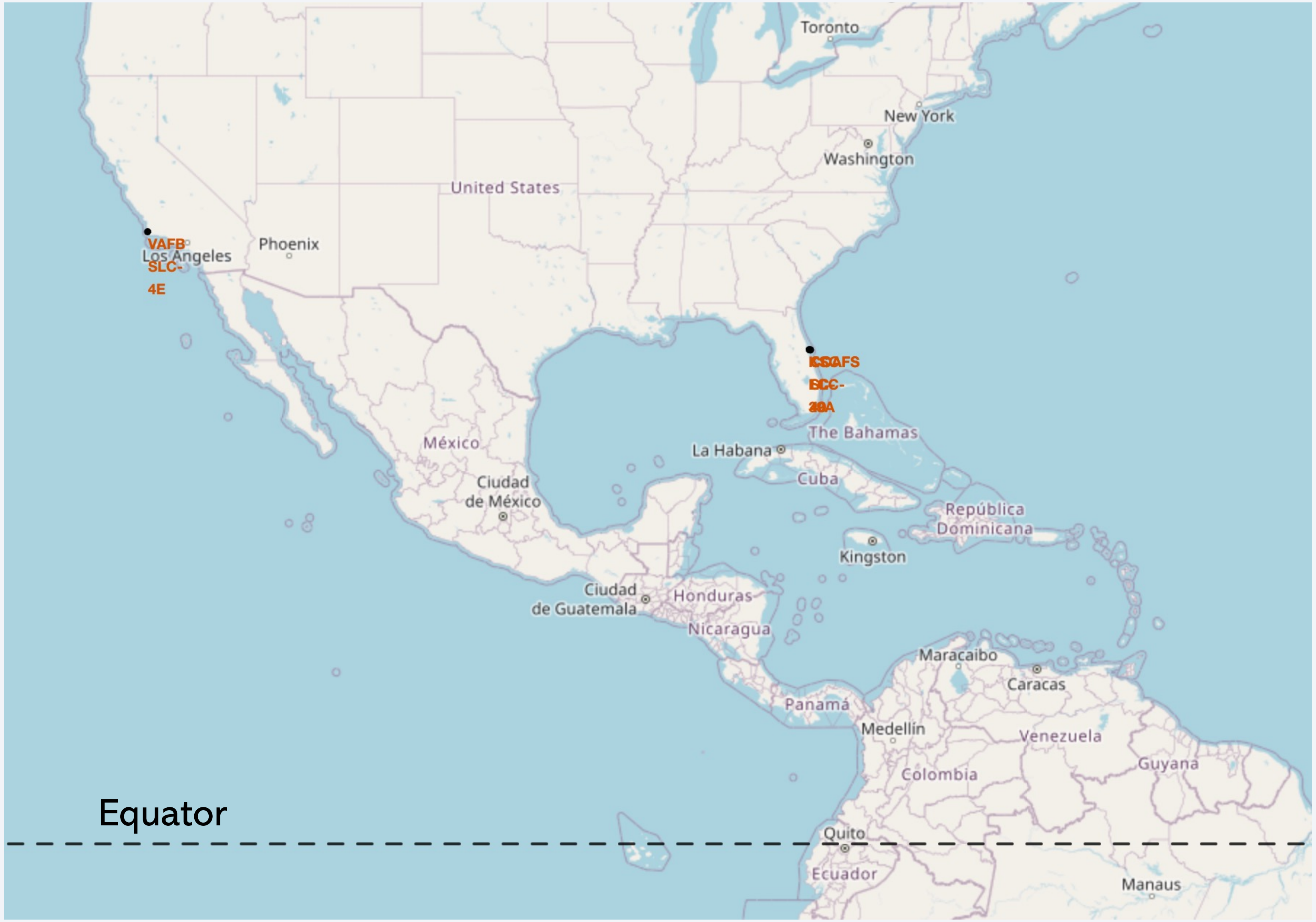
3. Landing Location Preferences:
- Drone ship was the most common landing target (11 attempts: 5 success, 5 failure, 1 precluded)
- Ground pad landings had the best success rate (3 successes, no failures)
- The ocean landing scenarios need further investigation, could be early tests or contingency scenarios (5 total attempts)
- Parachute landing attempts were relatively rare and unsuccessful (2 failures)

4. Evolution of Landing Technology:
- The variety of landing methods (parachute, drone ship, ground pad) suggests experimentation and evolution of landing technology
- The presence of ground pad successes and drone ship successes indicates SpaceX developed reliable landing capabilities for both scenarios

Section 3

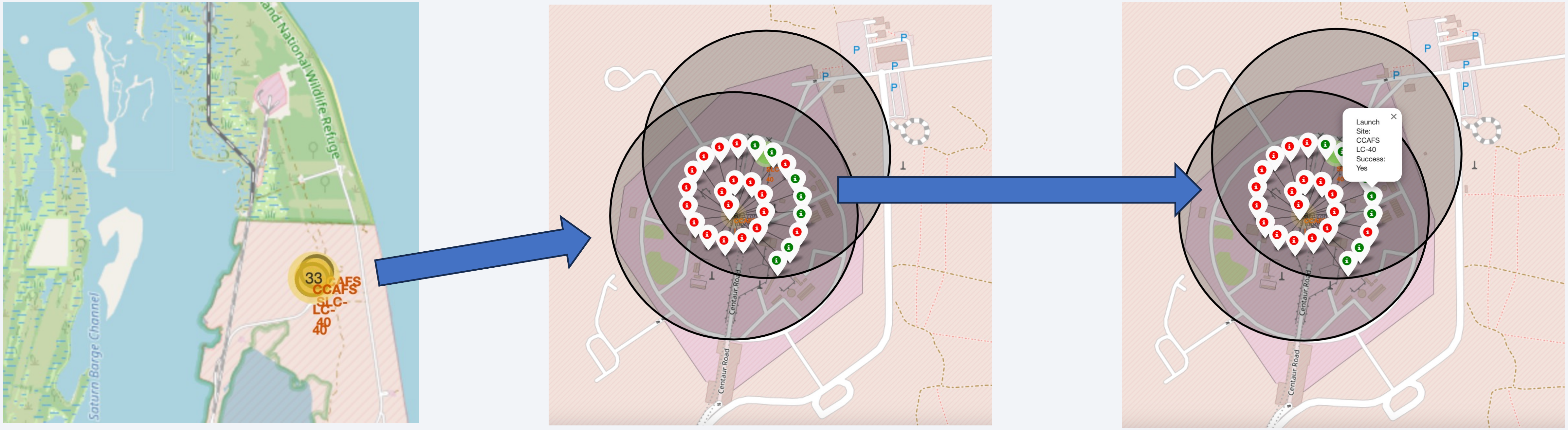
Launch Sites Proximities Analysis

Launch Site Locations



It can be seen from the global Folium map that the launch sites are located on the western and eastern coasts of the USA, as close to the Equator as it is possible to get within the borders of the USA

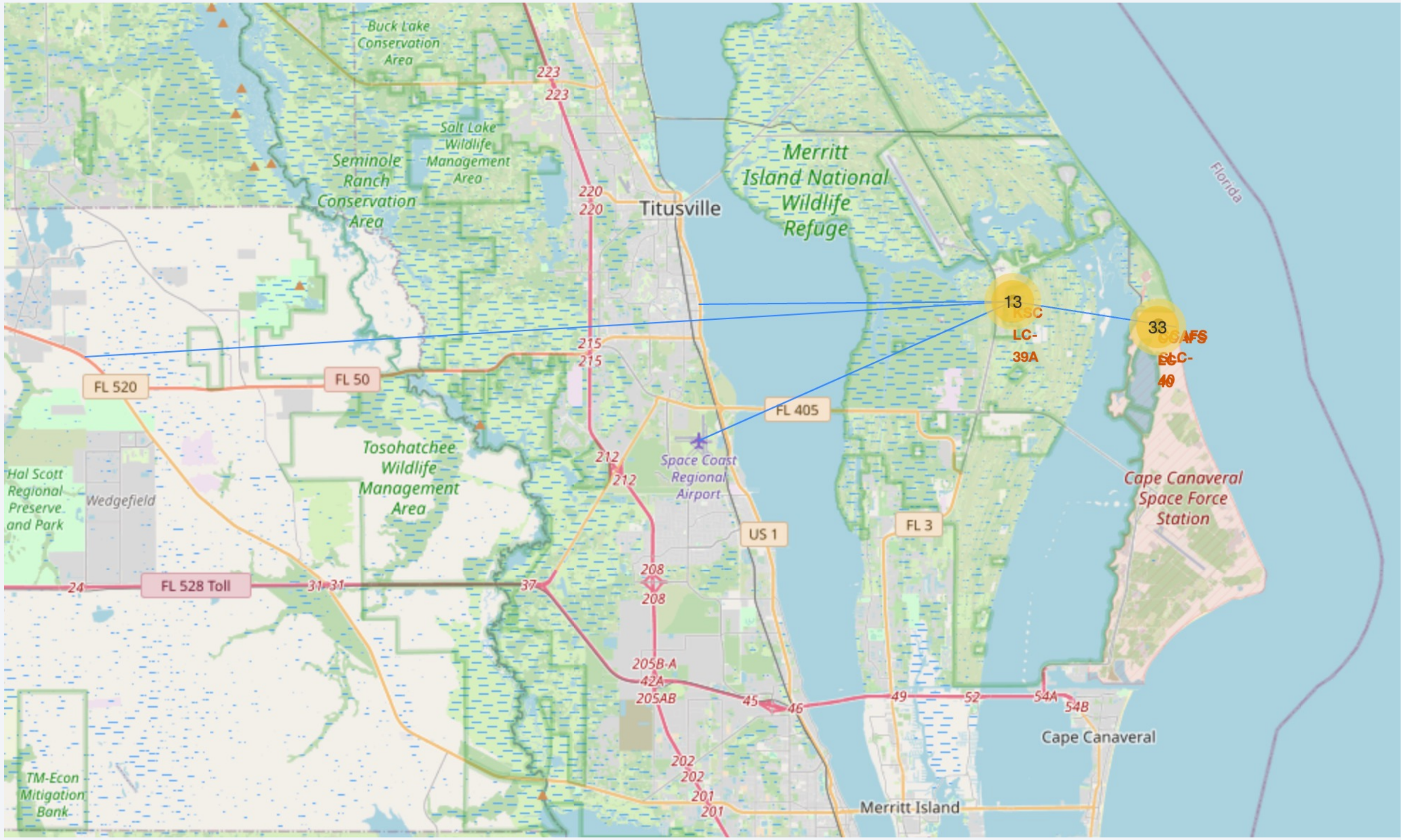
Colour-Coded Launch Outcomes on Folium Map



The interactive Folium map employs marker clustering to visualise launch outcomes across different sites. Launches are colour-coded, with green markers indicating successful missions and red markers representing failed attempts. Users can adjust the map's zoom level to view either a broad overview or detailed site-specific information. Each marker contains additional launch details that can be accessed with a click.

This visualisation method effectively highlights the varying success rates across launch facilities, with Kennedy Space Center's Launch Complex 39A demonstrating notably strong performance in booster landing operations.

Proximities of a Selected Launch Site



Calculated distances from Kennedy Space Center's Launch Complex 39A to nearby geographical features using a trigonometric function and displayed the measurements on a Folium map using polylines. The analysis revealed the following distances:

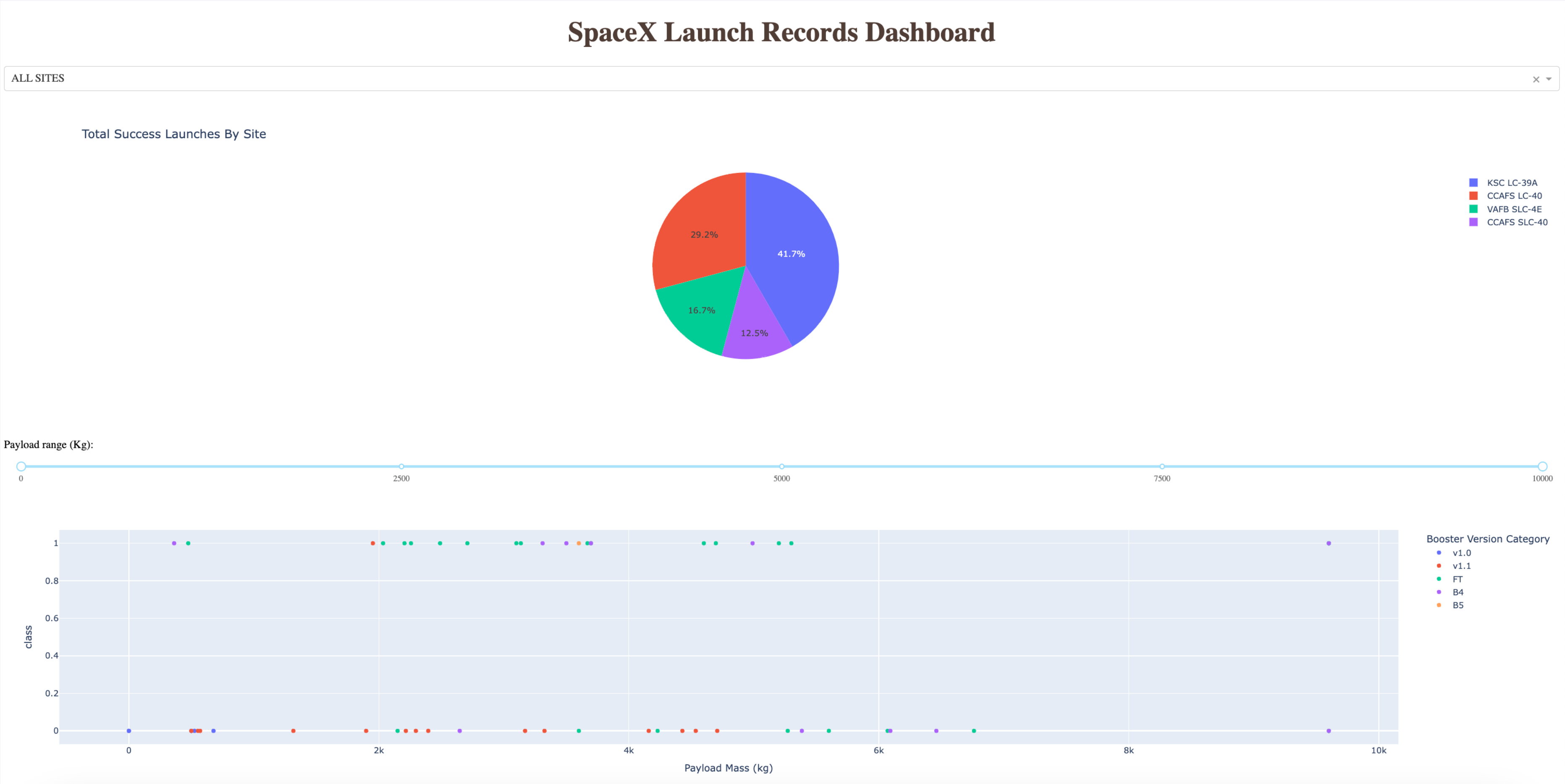
- Coastline: 7.8 km
- Highway: 14.8 km
- Airport: 16.2 km
- City: 43.9 km

These measurements demonstrate that launch sites are strategically positioned at safe distances from critical infrastructure and populated areas to mitigate risks associated with potential launch failures or accidents.

Section 4

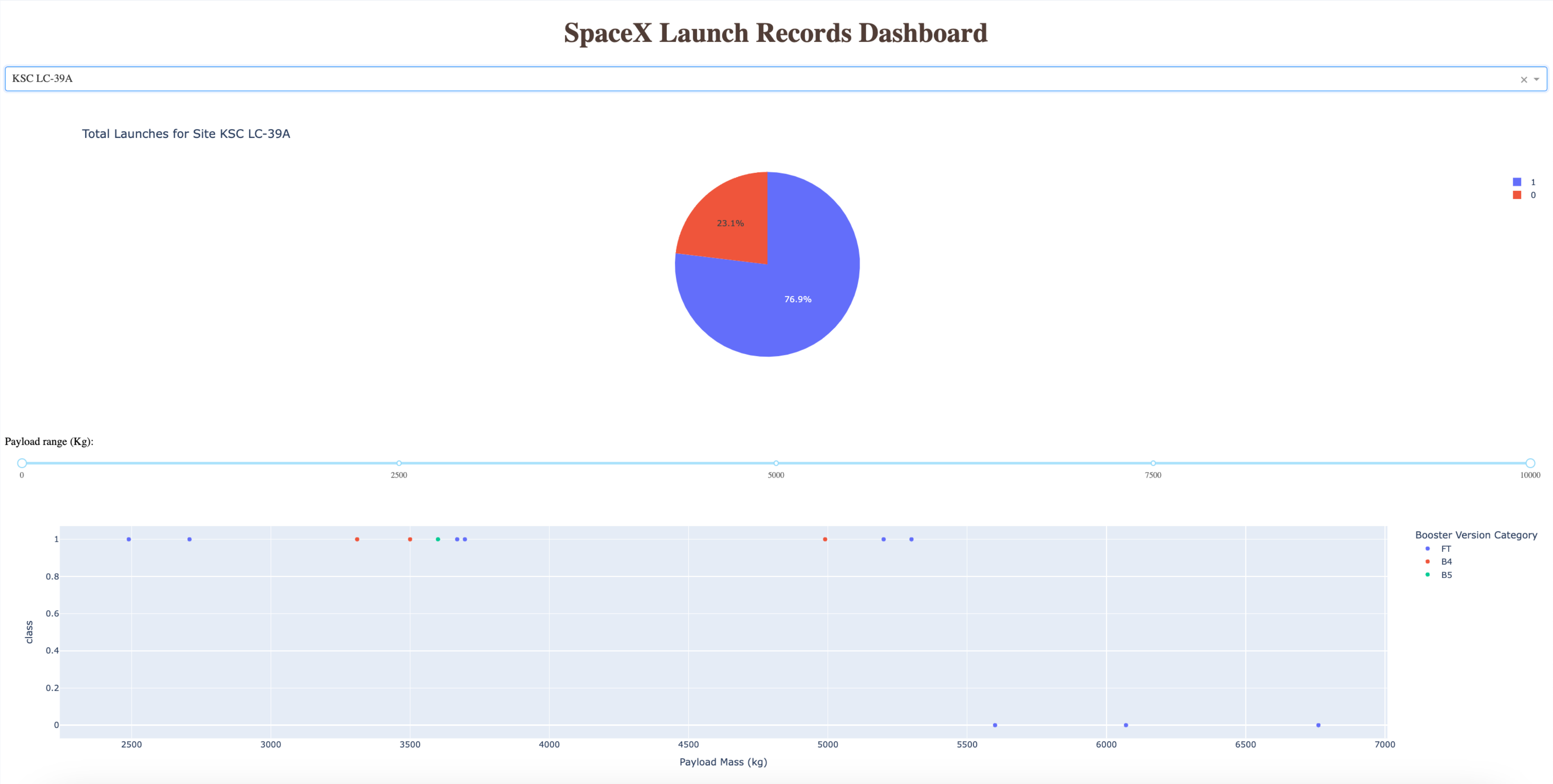
Build a Dashboard with Plotly Dash

Plotly Dashboard: Distribution of Successful Launches (All Sites)



The pie chart shows the distribution of successful SpaceX launches across the four different launch sites. KSC LC-39A accounts for the largest portion of successful launches (41.7%), followed by CCAFS LC-40 (29.2%).

Launch Site With The Highest Success Rate



Using the interactive pie chart to display the launch success rates of the different sites, it was easy to determine that Kennedy Space Center's Launch Complex 39A had the highest success rate at 77%

Payload vs. Launch Outcome

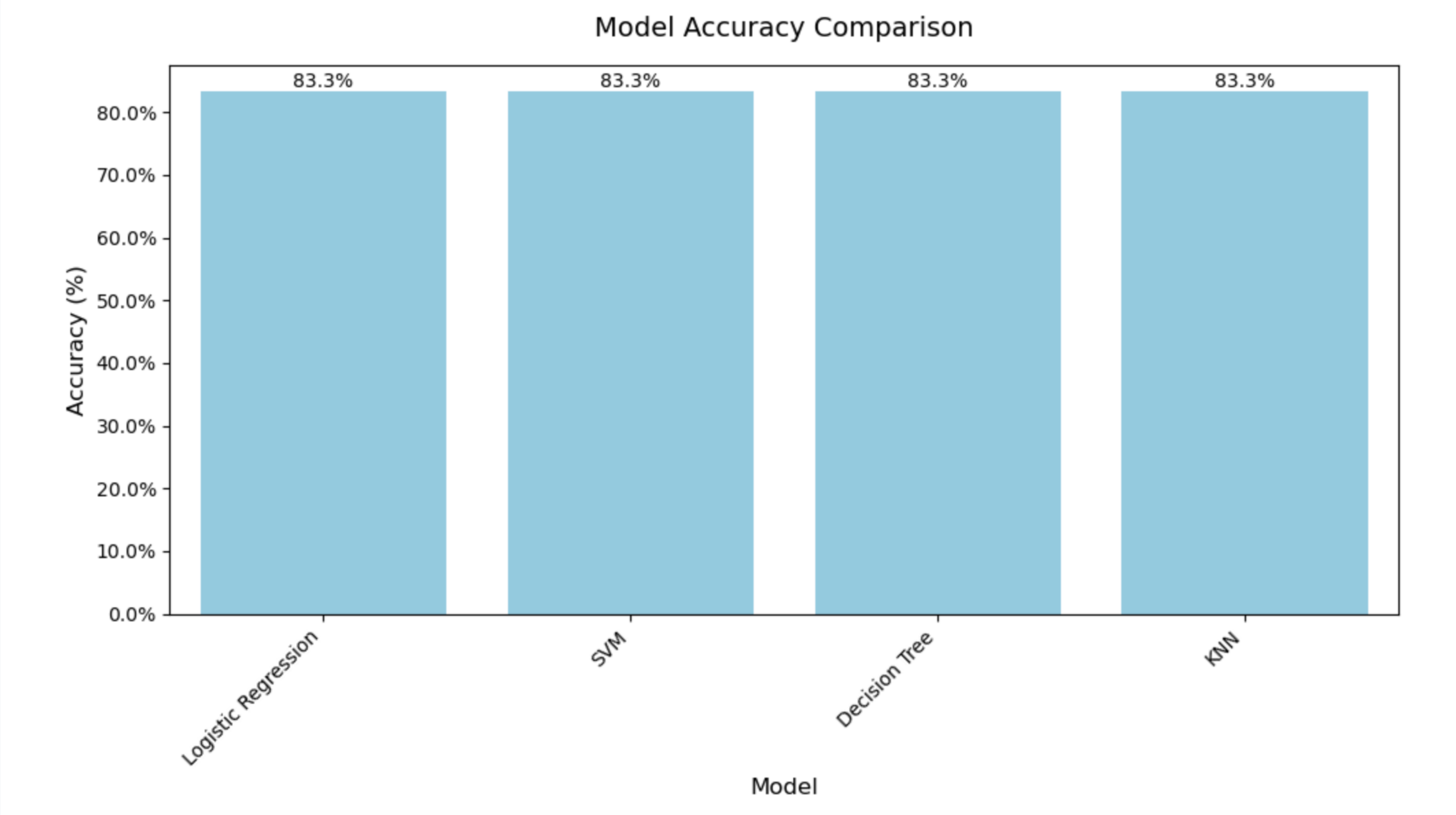
Using the interactive scatterplot display it was possible to determine that:

- Booster Version Category B4 had a 50% success rate in carrying the heaviest payload (9,600 kg)
- Payloads < 5,300 kg had the highest booster landing success rate, with the FT booster being the most successful in this range
- As the payload mass increases, the booster version category generally increases (from v1.0 to v1.1 to FT to B4 to B5), indicating the use of more powerful boosters for heavier payloads
- There is some overlap between the booster version categories, suggesting a continuous relationship between payload mass and booster version rather than strictly discrete categories

Section 5

Predictive Analysis (Classification)

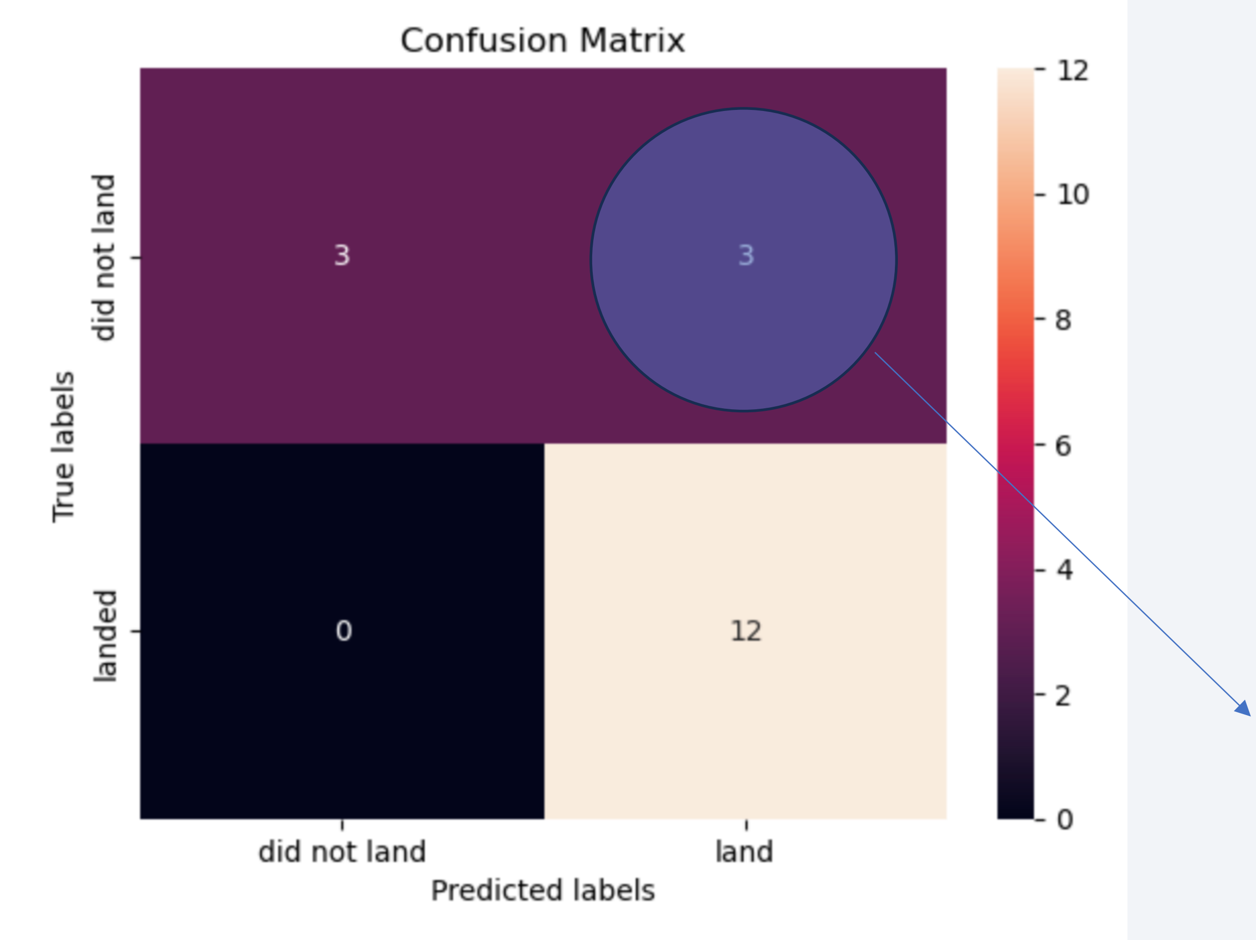
Classification Accuracy



It can be seen from the bar chart that all the models had the same test accuracy result of 83.3%

This counterintuitive result may be accounted for by the small test set (n=18), or by the fact that the problem is relatively straightforward so that all the models were able to learn the main patterns easily

Confusion Matrices



All the models generated the same confusion matrix

All the models were able to distinguish between the different classes, and all the models correctly predicted when missions had landed

However, all the models had a problem with false positives: in 3 cases the models predicted that a first stage booster would land when it didn't

Conclusions

SpaceX offers rocket launches at $62m compared to competitors charging over $165m, primarily due to their ability to reuse the first stage rocket booster

In order for SpaceY to learn from SpaceX's experience and bid successfully for launches against SpaceX, the following learning points can serve as a guide:

- Use launch site KSC LC-39A (near Cape Canaveral) if possible, as this site has a launch success rate of 77%
- Payloads of less than 5,300 kg were more successful than those of greater weight
- Orbits should be selected based on the payload carried (ISS missions cluster around 2,000-3,000 kg while VLEO missions appear successful with heavier payloads around 15,000 kg)
- The orbits with the most successful missions were ES-11, GEO, HEO and SSO
- Ground pad landings have been the most successful (100% with n=3), but drone ship landings may continue to improve (currently 45% with n=11)
- The classification models developed in this project can be used to predict successful stage 1 rocket landings with an accuracy of 83%

Appendix

- A7: Alternative Plotly Dash application recoded to be used directly in Jupyter Notebook, with interactive dashboard in external internet browser:

https://github.com/avanheerden/capstone/blob/main/A7-SpaceX_dashboard.ipynb

Acknowledgements

- With many thanks to Joseph Santarcangelo and the rest of the team at IBM for creating the course and project materials

- Visual assets: Microsoft Stock Image Library

Thank you!